2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Recognition of 6 DOF Rigid Body Motion Trajectories using a Coordinate-Free Representation

Joris De Schutter, Enrico Di Lello, Jochem F.M. De Schutter, Roel Matthysen, Tuur Benoit, and Tinne De Laet

***Abstract*— This paper presents an approach to recognize 6 DOF rigid body motion trajectories (3D translation + rotation), such as the 6 DOF motion trajectory of an object manipulated by a human. As a first step in the recognition process, 3D measured position trajectories of arbitrary and uncalibrated points attached to the rigid body are transformed to an invariant, coordinate-free representation of the rigid body motion trajectory. This invariant representation is independent of the reference frame in which the motion is observed, the chosen marker positions, the linear scale (magnitude) of the motion, the time scale and the velocity profile along the trajectory. Two classification algorithms which use the invariant representation as input are developed and tested experimentally: one approach based on a Dynamic Time Warping algorithm, and one based on Hidden Markov Models. Both approaches yield high recognition rates (up to 95 % and 91 %, respectively). The advantage of the invariant approach is that motion trajectories observed in different contexts (with different reference frames, marker positions, time scales, linear scales, velocity profiles) can be compared and averaged, which allows us to build models from multiple demonstrations observed in different contexts, and use these models to recognize similar motion trajectories in still different contexts.**

I. INTRODUCTION

In robotics there is a need to represent a rigid body motion trajectory in a way that facilitates recognition, classification and generalization to different contexts. In human intent estimation, recognition of the 6 DOF pose trajectory of a rigid object manipulated by a human, or of a human body segment such as the torso or the hand, may be an interesting alternative or complement to human gesture recognition based on joint angles [1]. For example this would be the case if the human is not sufficiently visible or not instrumented with markers. Similarly, in robot programming by robot guiding, the recorded motion of the manipulated object or robot hand may be an interesting alternative or complement to the recorded robot joint angles. For example, in [2], the 3D cartesian position (not the full 6DOF pose!) of the two robot hands is included in the observed variables. Finally, in robot programming by human demonstration, extraction of a motion primitive for the manipulated object from multiple demonstrations, possibly observed in different contexts, requires a representation of the trajectories of the manipulated object which allows us to compare and average them.

Recently, a coordinate-free representation of 6 DOF rigid body motion trajectories was proposed [3]. This representation models the intrinsic differential geometric properties of the trajectory, and is invariant with respect to the reference frame in which the motion is recorded, the reference point on the rigid body chosen to express the translational components of the motion, the velocity profile along the trajectory, as well as the time scale and linear or angular scale of the motion. These invariance properties are highly desirable in view of the envisioned applications mentioned above, because they allow us *1)* to learn models for motion primitives from multiple demonstrations in different contexts and environments, and *2)* to use these models in still different contexts and environments. To exploit these invariance properties, measured motions are transformed to their invariant representation *before* applying algorithms for recognition, classification, clustering, comparison, averaging, etc., to get rid of the dependency on the reference frame, reference point, parametrization, time scale and linear or angular scale, and hence to reduce the search space for these algorithms.

The invariant representation used in this paper is a generalization of the invariant signature that was proposed in [4] for effective motion trajectory recognition in two respects, because it is not limited to the *planar* motion of a single *point*, but applies to the *spatial* motion of a *rigid body*. Similarly, the view-invariant representation of trajectories introduced in [5] only deals with spatial trajectories of a *point*, not of a rigid body, hence without considering orientation. The invariant representation used here is also quite different from the invariants used in vision, such as the space-time invariants for 3D motions from projective cameras [6]. A major difference is that the projective invariants in [6] are calculated for 3D position trajectories of individual *points*, not 6 DOF pose trajectories of rigid bodies. Hence, to compare different observations it is required that each time the motion of the *same* points is observed. This is not the case for the invariant representation used here: different points attached to the rigid body may be chosen to measure the motion of the rigid body in the different observations. This is an important advantage in view of the generalization abilities discussed above. On the other hand, the projective invariants allow us to compute the invariants directly from the image coordinates, whereas the invariants used here assume that

Jochem De Schutter, Roel Matthysen and Tuur Benoit are engineering students at K.U.Leuven, Belgium, {jochem.deschutter,roel.matthysen,tuur.benoit}@student.kuleuven.be

All other authors are with the Department of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300, Box 02420, B-3001 Leuven-Heverlee, Belgium, {joris.deschutter,tinne.delaet,enrico.dilello}@mech.kuleuven.be

an explicit 3D reconstruction of measured points attached to the rigid body is available, for example from multiple camera images.

This paper focuses on motion recognition of a single rigid body in 6-dimensional space (3D translation and rotation). A person manipulates a rigid object while its motion is observed. Different motions are executed, each with multiple repetitions (trials). **The main contribution of the paper is to introduce an invariant representation as a starting point for motion recognition algorithms. The invariant representation reduces the variability of the motion data *at the source*. As a result, very good recognition results can be obtained even with very basic classification algorithms.** By way of proof-of-concept we compare the use of two basic algorithms, Dynamic Time Warping (although we do not warp the time, but another variable) and Hidden Markov Models, to build a model for each of the motions and to perform recognition/classification experiments. Recognition rates up to 95% and 91% are obtained for the Dynamic Time Warping-based approach and the Hidden Markov Models-based approach, respectively.

The paper is organized as follows. Section II introduces the invariant representation used in this paper. Section III describes the experimental data collection and preprocessing steps. Sections IV and V explain the Dynamic Time Warping (DTW)-based and Hidden Markov Modeling (HMM)-based modeling and recognition approaches, respectively. Section VI presents the experimental results, while Section VII discusses these results, states the conclusions and points to future work.

II. THEORETICAL BACKGROUND

This section briefly reviews the invariant representation of 6 DOF rigid body pose trajectories introduced in [3][1].

A. Time-Based Invariants

Any rigid body motion is characterized instantaneously to the first order by the instantaneous screw axis (ISA) [7], [8]. The rotational velocity about and translational velocity along the ISA, denoted by scalars ω_1 and v_1, respectively, are two invariants of the rigid body motion. Four other invariants, two rotational and two translational velocities, model the spatial motion of the ISA. These invariants are denoted by scalars ω_2 and ω_3, and v_2 and v_3, respectively. While ω_2 and v_2 represent the instantaneous rotational and translational velocity of the ISA, ω_3 and v_3 represent the instantaneous rotational and translational velocity of the common normal between two instances of the ISA at infinitesimally separated time instants. For a general rigid body motion the six invariants are a function of time, but special motions exhibit particular invariant functions, which can be used to recognize them. For example, a planar motion is characterized by $\omega_2(t) \equiv 0$ and $v_1(t) \equiv 0$. A hinge motion is further characterized by $v_2(t) \equiv 0$. See [3] for the invariant *signature* of other special motions.

[1]This paper only contains the theoretical foundations of the invariant representation, and does not include any experiments.

In [3] analytic formulas are presented to obtain the *time-based invariants* $\omega_1(t)$, $\omega_2(t), \omega_3(t), v_1(t), v_2(t)$ and $v_3(t)$ starting from the twist of the rigid body, given as $\mathbf{t} = (\boldsymbol{\omega}^T \ \boldsymbol{v}^T)^T$, and its first and second time derivatives. The coordinates of $\boldsymbol{\omega}$ and $\boldsymbol{v}$ are expressed in the world reference frame if we are interested in the *absolute* motion of the rigid body, or in any other reference frame attached to a second body if we are interested in the *relative* motion with respect to this second body. On the other hand, the reference point used to express the translational velocity $\boldsymbol{v}$ is the point attached to the rigid body that instantaneously coincides with the origin of the chosen reference frame. These analytic formulas are listed in the appendix.

The transformation from the rigid body twist to the invariants exhibits singularities when ω_1 or ω_2 is zero, that is, in the case of a pure translation or an ISA with constant orientation, respectively. In these cases some of the invariants are not defined, see [3] for details.

B. Geometric Invariants

To eliminate the influences of the time scale and of the velocity profile with which the motion is executed, and hence to only retain the geometric properties of the motion, the invariants are expressed in terms of a geometric *degree of advancement*, $\xi(t)$, instead of time. Its time derivative is the *rate of advancement*, $\dot{\xi}(t)$. In order to be applicable not only to a general motion but also to the special cases of pure rotation ($v_1(t) \equiv 0$) and pure translation ($\omega_1(t) \equiv 0$), the rate of advancement is defined as:

$$\dot{\xi}(t) = w\frac{|\omega_1(t)|}{\Theta_s} + (1-w)\frac{|v_1(t)|}{L_s} \ , \tag{1}$$

where Θ_s and L_s represent user-defined scaling factors with dimensions of angle and length, respectively, and $0 \leq w \leq 1$ is a user-defined dimensionless weight. Hence the degree of of advancement $\xi(t)$ is dimensionless.

Once the scales and the weight have been fixed, dividing the time-based invariants by the rate of advancement (1), inverting $\xi(t)$, and substituting t by ξ, results in six *geometric invariants* that contain the geometry of the rigid body motion without the effect of time. The geometric invariants are denoted by capital letters, $\Omega_i(\xi)$ and $V_i(\xi)$, with $i = 1, 2, 3$:

$$\Omega_i(\xi) = \frac{\omega_i(t(\xi))}{\dot{\xi}(t(\xi))} \ ; \ V_i(\xi) = \frac{v_i(t(\xi))}{\dot{\xi}(t(\xi))} \ . \tag{2}$$

Only five geometric invariants are independent, since $\Omega_1(\xi)$ and $V_1(\xi)$ are constrained by (1), which after division by $\dot{\xi}(t)$ becomes:

$$1 = w\frac{|\Omega_1(\xi)|}{\Theta_s} + (1-w)\frac{|V_1(\xi)|}{L_s} \ . \tag{3}$$

To completely describe the motion, the five geometric invariants have to be supplemented with $\xi(t)$, which contains the temporal information.

C. Dimensionless Geometric Invariants

To compare motions with different angular or linear amplitudes, the degree of advancement can be scaled to 1 by selecting in (1) $\Theta_s = \Theta$ and $L_s = L$ defined as:

$$\Theta = \int_{t_0}^{t_f} |\omega_1| dt \ ; \ L = \int_{t_0}^{t_f} |v_1| dt \ , \tag{4}$$

where t_0 and t_f represent the start and end time of the motion, respectively. This procedure yields the normalized degree of advancement $\bar{\xi}(t)$. This procedure however breaks down if the entire motion consists of a pure translation or a pure rotation, because $\Theta = 0$ or $L = 0$, respectively. In such case, which can be detected, w is set equal to 1 or 0 in (1), respectively. Furthermore, dividing $\omega_i(t)$ and $v_i(t)$ by $\dot{\bar{\xi}}$ and dividing them by Θ and L, respectively, inverting $\bar{\xi}(t)$, and substituting t by $\bar{\xi}$ results in six *dimensionless geometric invariants*, denoted by $\bar{\Omega}_i(\bar{\xi})$ or $\bar{V}_i(\bar{\xi})$:

$$\bar{\Omega}_i(\bar{\xi}) = \frac{\omega_i(t(\bar{\xi}))}{\dot{\bar{\xi}}(t(\bar{\xi}))\Theta} \ ; \ \bar{V}_i(\bar{\xi}) = \frac{v_i(t(\bar{\xi}))}{\dot{\bar{\xi}}(t(\bar{\xi}))L} \ . \tag{5}$$

Using the definitions of $\bar{\Omega}_i(\bar{\xi})$ and $\bar{V}_i(\bar{\xi})$, constraint (3) reduces to:

$$1 = w|\bar{\Omega}_1(\bar{\xi})| + (1-w)|\bar{V}_1(\bar{\xi})| \ , \tag{6}$$

while the temporal invariant is conveniently written as a function of dimensionless time: $\bar{\xi}(\bar{t})$, where $\bar{t} = \frac{t}{t_f - t_0}$. Notice that $0 \leq \bar{\xi}(\bar{t}) \leq 1$ and $0 \leq \bar{t} \leq 1$.

III. DATA COLLECTION AND PREPROCESSING

A. Measurement Set-Up

A person manipulates a rigid object, while the motion of LED markers attached to the object is recorded. The marker positions are chosen arbitrarily and do not need to be calibrated with respect to the object reference frame. The measurement system consists of a Krypton K600 camera system from NIKON Metrology, figure 1 (right), which comprises three calibrated line cameras. Each LED marker is fired separately, and, if the LED is visible in the camera system, its position is recorded by the three line cameras, after which the 3D position of the LED marker is reconstructed with respect to the reference frame attached to the camera system. The camera system is fast and highly accurate, taking measurements at 100 Hz or more, depending on the number of markers used, with a volumetric accuracy of 90 μm. To increase the probability that at least three LED markers are visible by the camera system throughout an entire motion sequence a rigid object with multiple faces is used, figure 1 (left), to which up to 12 LED markers can be attached.

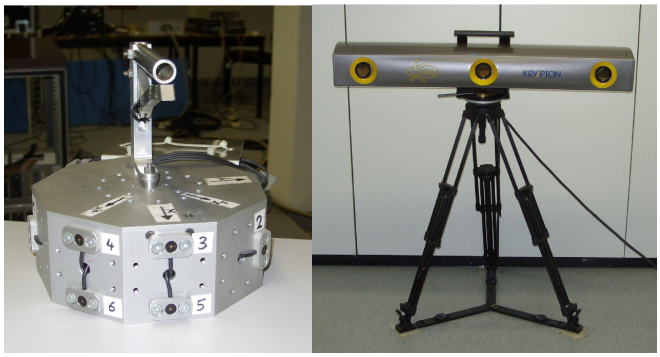

Fig. 1. *Left:* object with handle on top and with multiple faces to which up to 12 LED markers can be attached; *right:* Krypton K600 camera system.

B. Data collection

A person was asked to perform 20 demonstrations of nine different but freely selected motions involving both rotation and translation of the object, yielding a total set of 180 recorded motion trials. Eleven motion trials were rejected, because less than three markers were visible throughout the entire motion sequence, leaving a set of 169 valid motion trials. A video clip of one demonstration is available at `http://people.mech.kuleuven.be/~jdeschut/icra2011`. For each of the nine motions, the person was asked to perform similar, but not exactly identical demonstrations. In between demonstrations the person was allowed to change the starting position and orientation with respect to the camera system. As a result, different sets of LED-markers were visible by the camera system for the different demonstrations. We will make this data set available online as 'data set1'. Figure 2 shows a 3D view of a sample trial of each of the nine motions.

C. Preprocessing Steps

The calculation of the invariant representation for each motion trial requires the twist of the object and its first and second time derivatives as input. Obtaining these quantities from the 3D measured marker positions $\boldsymbol{p}_i(t)$ requires a number of preprocessing steps.

First, a Kalman smoothing algorithm [9] is applied to each coordinate of $\boldsymbol{p}_i$ separately to obtain the first three time derivatives (velocity, acceleration and jerk) in a numerically stable way. The advantage of a Kalman smoother over a Kalman filter is that there is no time delay between the measured and smoothed signal. The disadvantage is that the signals can only be processed after completion of the entire motion sequence.

Next, the twist of the object is calculated. Several authors have developed efficient algorithms to compute rigid body velocity and/or acceleration descriptors based on position, velocity, and acceleration data for individual points of the moving body [10], [11], [12], [13], [14]. Fenton and Willgoss [15] compare five methods for determining the twist of a rigid body from position and velocity data of individual points attached to it. We have however used another method which can easily be applied to an arbitrary number of visible markers ($m \geq 3$), and which is easily extended to obtain also the time derivatives of the twist. While the twist components $\boldsymbol{\omega}$ and $\boldsymbol{v}$ are solved from the overdetermined set of linear equations

$$\dot{\boldsymbol{p}}_i = \boldsymbol{v} + \boldsymbol{\omega} \times \boldsymbol{p}_i \ , \ i = 1, \ldots, m, \tag{7}$$

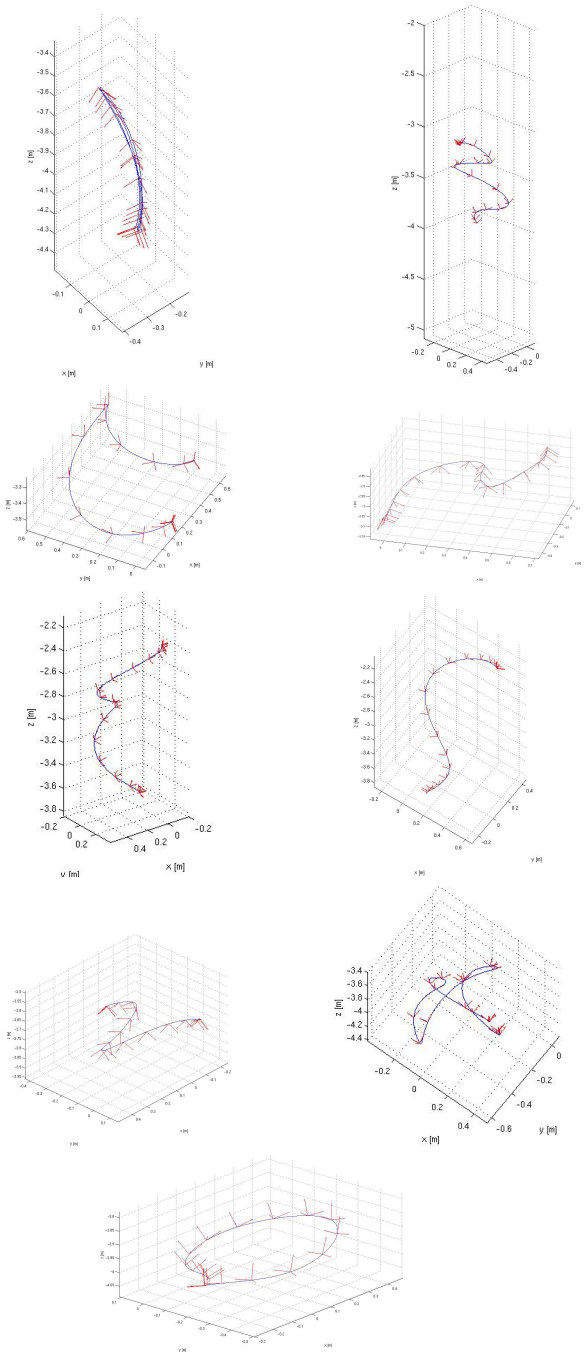

Fig. 2. 3D view of a sample trial of each of the nine motions, showing both translation and rotation.

their first and second time derivatives are subsequently solved from the first and second time derivative of (7), respectively.

Subsequently, the dimensionless geometric invariants are calculated for each motion trial using the procedure outlined in Section II. While $\bar{\Omega}_1$ and $\bar{V}_1$ can only take values between -2 and $+2$ due to eq. (6), $\bar{\Omega}_2$, $\bar{\Omega}_3$, $\bar{V}_2$ and $\bar{V}_3$ can take values between $-\infty$ and $+\infty$. Therefore we scaled these dimensionless geometric invariants to take values between -1 and $+1$ by defining, for $i = 2, \ldots, 3$:

$$\bar{\Omega}_i^* = \frac{\bar{\Omega}_i}{s_{\Omega_i} + |\bar{\Omega}_{V_i}|} \; ; \; \bar{V}_i^* = \frac{\bar{V}_i}{s_{V_i} + |\bar{V}_i|} \; , \tag{8}$$

where the asterisk denotes 'scaled', and s_{Ω_i} and s_{V_i} denote user-selected constants.

D. Assumptions, choices and parameters

In our experiments the measurement frequency was 200 Hz. We used a fifth-order Kalman smoother, assuming motions with constant derivative of jerk. We assumed gaussian process and measurement noise, and chose the ratio of their variances as $6.25 \times 10^{10} s^{-8}$. The weight in eq. (1) was chosen $w = 0.5$. Since $\bar{\Omega}_1$ contains the same information as $\bar{V}_1$ due to constraint (6), we did not use $\bar{\Omega}_1$ in the modeling and recognition procedure outlined in Sections IV and V. Furthermore, in (8) $s_{\Omega_2} = 10 rad/s$, $s_{V_2} = 20 m/s$, $s_{\Omega_3} = 30 rad/s$ and $s_{V_3} = 30 m/s$.

IV. DTW-BASED MODELING AND RECOGNITION

A. Background

Dynamic time warping[16] is an algorithm for measuring similarity between two sequences that may vary in time or speed. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. The similarity is expressed in terms of the DTW distance. In this paper DTW is applied to the dimensionless geometric invariants (or their scaled versions), hence the invariant functions are warped non-linearly in the dimension of the dimensionless degree of advancement, $\bar{\xi}$, rather than the dimension of time. For calculating the DTW distance we used a rather basic algorithm[2].

B. Model construction

This section shows how to construct a model for each motion class c_j, with $j = 1, \ldots, 9$, and for each invariant i from a number n_m of training motion trials. In the DTW-approach we ignored invariants $\bar{\Omega}_3^*$ and $\bar{V}_3^*$, because we found that they did not improve the recognition rate[3]. Hence we focused on invariants $\bar{V}_1(\bar{\xi})$, $\bar{\Omega}_2^*(\bar{\xi})$ and $\bar{V}_2^*(\bar{\xi})$, and hence 27 models are constructed (9 motions $\times$ 3 invariants). Each model consists of an averaged invariant function, together with a probability density function (pdf) $p(DTW_{ij}|c_j)$ that represents the probability of the DTW distance between the i-th averaged invariant function of motion class j and the corresponding invariant function of a trial of the same motion class. The procedure is briefly outlined below.

For each of the nine motions, n_m training motion trials used to build the model are selected randomly from the available trials. For each motion class c_j and for each invariant i a model is then built as follows. First, for each training trial $k = 1, \ldots, n_m$ the average DTW distance to all other training trials, $DTW_{av,k}$ is calculated (for convenience we

[2]given at $http://en.wikipedia.org/wiki/Dynamic\_time\_warping$

[3]$\bar{\Omega}_3^*$ and $\bar{V}_3^*$ are based on the third derivative of the marker positions, hence they are more difficult to estimate reliably by the Kalman smoother.

omit subscripts i and j in the remainder of this subsection). An exponential pdf is fitted through these $DTW_{av,k}$-values:

$$p(DTW_{av}|\lambda) = \lambda \exp^{-\lambda DTW_{av}}, \quad (9)$$

where $\lambda^{-1} = \frac{1}{n_m}\sum_{k=1}^{n_m} DTW_{av,k}$. This pdf is used to weight the contribution of the n_m training trials to the model. Building the model proceeds in an iterative way, starting with the two training trials with the lowest $DTW_{av,k}$, because they have the highest similarity with the other training trials. The two training trials are fused together by calculating a weighted average of corresponding $\bar{\xi}$-values in the two trials as identified in their mutual DTW analysis. The iteration proceeds by repeating this procedure between the calculated average and the other training trials, in order of increasing $DTW_{av,k}$, hence decreasing similarity. Figure 3 illustrates the result of this procedure for one motion class and one invariant.

Next, the DTW distance between each of the n_m training motion trials and the model is calculated and an exponential pdf, $p(DTW_{ij}|c_j)$, is fitted through these values.

C. Recognition

All remaining motion trials, i.e. $169 - 9n_m$, are used in recognition/classification experiments. A rejection class c_{rej} is defined [17], to classify the trials that have a bad fit with the models of each of the nine motion classes, hence trials that are not recognized. Furthermore, the nine motion classes c_j are assigned an equal a priori probability $p(c_j)$.

The classification of a motion trial then proceeds as follows. First, for each motion class j and each invariant i the DTW distances DTW_{ij} between the motion trial and each of the motion models is calculated. Next, the likelihoods $p(DTW_{ij}|c_j)$ are calculated. These likelihoods are combined for the three invariants $i = 1, 2, 3$ of each motion class c_j:

$$p(DTW_{ij}, i = 1, 2, 3|c_j) = \prod_{i=1}^{3} p(DTW_{ij}|c_j), \quad (10)$$

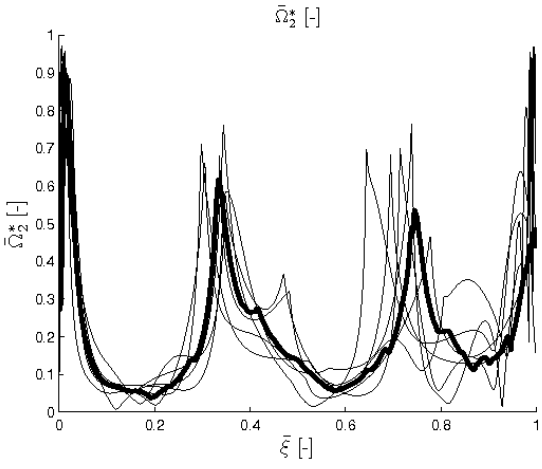

Fig. 3. Construction of the model for $\bar{\Omega}_2(\bar{\xi})$ for one motion, with $n_m = 6$. The thin lines correspond to the six individual trials; the thick line represents the averaged invariant function.

which is based on the assumption that the DTW distances corresponding to the three invariants are independent, given the motion class c_j.

Then, the total data density of the known classes is calculated as [17]:

$$p(DTW^*) = \sum_{j=1}^{9} p(DTW_{ij}, i = 1, 2, 3|c_j)p(c_j). \quad (11)$$

If $p(DTW^*) < \theta$, where θ is a rejection threshold, the trial is assigned to the rejection class. Otherwise the trial is assigned to the motion class with the largest a posteriori probability $p(DTW_{ij}, i = 1, 2, 3|c_j)p(c_j)$.

D. Assumptions, choices and parameters

The invariants are represented by 500 data points[4]. The DTW algorithm needs a window parameter that limits the search for the best fit. This window parameter was chosen 10% of the interval of $\bar{\xi}$, i.e. equal to 0.1. We chose to work with exponential distributions; other choices are possible. The threshold θ was chosen 0.46.

V. HMM-BASED MODELING AND RECOGNITION

A. Background

The *Hidden Markov Model* is a statistical signal model widely used in speech recognition, natural language modeling, on-line handwriting recognition, and for the analysis of biological sequences such as proteins and DNA [18].

The HMM can be seen as a stochastic finite state automaton, where each state emits an observation. More in detail, at each time t the observations Y_t are considered a probabilistic function of the state X_t . The state X_t is represented by a discrete random variable $\in 1..K$, evolving according to a stochastic process that is not observable (i.e. *hidden*), and can only be observed through the produced sequence of observations [19]. An HMM is defined by the following set λ of model parameters:

- an initial state distribution $\Pi = [\pi_1, \pi_2, \ldots, \pi_n]$, where $\pi(i) = P(X_1 = i)$. π is represented by a multinomial distribution.
- a transition model, represented by the stochastic matrix A where $A(i, j) = P(X_t = j|X_{t-1} = i)$, modeling the evolution of the unobservable state. It is usually characterized by a conditional multinomial distribution.
- an observation model, which defines the probabilities $P(Y_t|X_t)$. If the observations are discrete symbols then $Y_t \in \{1...L\}$ and the observation model is represented as a matrix $B(i, j) = P(Y_t = j|X_t = i)$. If the observations are continuous feature vectors then $Y_t \in R^L$ and many probabilistic models can be adopted to represent $P(Y_t|X_t)$.

[4]Later we obtained comparable results with only 200 data points.

B. Model construction

In our approach, we define an HMM for each of the nine different motion classes. The construction of motion models consists of finding the values of model parameters λ that maximize the likelihood function $P(Y|\lambda)$ of the provided training motion trials given the model parameters.

For an HMM the estimation of the maximum likelihood parameters is usually performed through the well-known Baum-Welch [20] algorithm, which belongs to the class of the EM algorithms. As in the case of the DTW-based classification approach, the n_m motion trials used to train the HMM are selected randomly among all the available trials. As opposed to the DTW approach, we verified experimentally that the HMM approach obtains better classification results when considering all five independent invariants.

Again, the invariant functions are defined over the dimensionless degree of advancement, so the observation vector is given by : $Y_{\bar{\xi}} = [\bar{V}_1(\bar{\xi}), \bar{\Omega}_2^*(\bar{\xi}), \bar{V}_2^*(\bar{\xi}), \bar{\Omega}_3^*(\bar{\xi}), \bar{V}_3^*(\bar{\xi})]$.

The chosen observation model is a multivariate Gaussian distribution, so that $P(Y_{\bar{\xi}} = y|X_{\bar{\xi}} = i) = \mathcal{N}(y, \mu_i, \Sigma_i)$ where

$$\mathcal{N}(y;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{L}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)) \tag{12}$$

with μ and Σ representing the mean value and the covariance matrix of the Gaussian, respectively.

The training trials for each motion class are fed to the Baum-Welch algorithm which iterates until a local maximum of the likelihood function $P(Y|\lambda)$ is reached. The values obtained for the HMM parameters will be used for the recognition of previously unseen motion trials.

C. Recognition

For the classification of new observation sequences Y we adopt a *maximum a posteriori* (MAP) approach. We assign to each of the nine motion classes the same *a priori* probability $p(c_j)$. Given new observation sequence Y, its likelihood given a set of parameters $P(Y|\lambda)$ is computed for all the learned HMMs, using the standard *forward-backward* procedure [19]. The likelihood and the prior are combined using Bayes's rule. If the maximum MAP probability is below a threshold θ, the motion trial is rejected.

D. Assumptions, choices and parameters

The invariants are sub-sampled to 250 data points. The dimensionality K of the hidden space is a parameter that influences the ability of the HMM to discriminate between different motions, but its value is highly related to the amount of training data available. In our experiments, we obtained the best results with $K = 8$. The initial choice for the parameter Π is a uniform distribution, while the matrix A for the transition model, and the parameters μ and Σ for the observation model are initialized with random values. The rejection threshold θ is set to 0.3.

TABLE I

PERCENTAGE OF CORRECTLY RECOGNIZED TRIALS, WRONGLY RECOGNIZED TRIALS, AND REJECTED (UNRECOGNIZED) TRIALS FOR RECOGNITION EXPERIMENTS USING THE INVARIANT REPRESENTATION (OR FOR COMPARISON, BETWEEN BRACKETS, USING THE TWISTS).

	% correct	% wrong	% rejected
DTW-approach			
data set 1	95.2 (93.4)	2.6 (0.0)	2.2 (6.0)
data set 2	94.3 (46.5)	4.2 (44.1)	1.5 (9.4)
HMM-approach			
data set 1	91.0	7.3	1.7
data set 2	86.1	8.9	5.0

VI. RESULTS

For both the DTW-approach and the HMM-approach, 100 experiments were performed in which each time $n_m = 6$ training trials were selected randomly to construct the models, while the classification algorithms were applied to all 115 remaining motion trials. Table I displays the recognition results, averaged over the 100 experiments, and expressed in terms of the percentage of correctly recognized trials, wrongly recognized trials, and rejected (unrecognized) trials. The success rates of the DTW-based and HMM-based approaches are approximately 95% and 91%, respectively.

VII. DISCUSSION AND CONCLUSION

The main contribution of this paper is to introduce a coordinate-free, invariant representation for the recognition of rigid body motion trajectories. The advantage of using an invariant representation is that the variability of the motion data is reduced at the source. This moves the computational burden from complex recognition algorithms to a more involved preprocessing of the motion data. As a result, basic algorithms can be used for classification. By way of proof-of-concept, two classification approaches, one based on a Dynamic Time Warping algorithm and one based on Hidden Markov Models, which use the invariant representation as input, have been worked out and tested experimentally, yielding very high recognition rates. It should also be pointed out that *all* motion trials were used in the experiments. Although it was evident that the data set contained some outliers, not a single motion trial was ignored when selecting the training trials or performing the recognition.

To show the power of the invariant approach, we have repeated the experiments starting from a modified data set. In the modified data set, the 3D measured marker positions in each trial of data set 1, $\boldsymbol{p}_i(t)$, $i = 1, \ldots, m$, where m is the number of visible markers, were transformed according to:

$$[\boldsymbol{p'}_1(t')\ \boldsymbol{p'}_2(t') \ldots \boldsymbol{p'}_m(t')] = s\,[\mathbf{1}_{3\times3}\ \mathbf{0}_{3\times1}]\,\boldsymbol{T} \times \begin{bmatrix} \boldsymbol{p}_1(t) & \boldsymbol{p}_2(t) & \ldots & \boldsymbol{p}_m(t) \\ 1 & 1 & \ldots & 1 \end{bmatrix} \boldsymbol{A}\ . \tag{13}$$

Here $t' = f(t)$ and $f(t)$ is a monotonously increasing function representing a change of velocity profile along the trajectory and a change of time scale; s is a scalar representing a change of linear scale; $\boldsymbol{T}$ is a homogeneous

transformation matrix representing a change of reference frame; and $\boldsymbol{A}$ is an $m \times m$ linear transformation matrix where the entries in each column sum to 1, representing a change of marker positions on the rigid object. f, s, $\boldsymbol{T}$ and $\boldsymbol{A}$ are chosen randomly for each motion trial. Transformation (13) mimics that each individual motion trial originates from a different context. We will make the transformed data set available online as 'data set2'.

Owing to the properties of the invariants, corresponding motion trials in data sets 1 and 2 yield identical invariants, at least in theory. In practice, while a change of linear scale and a change of reference frame do not affect the invariants at all, a change of the marker positions and a change of velocity profile do affect the invariants. Since (7) represents an overdetermined set of equations, its solution depends on the weight of each equation, hence on the marker positions $\boldsymbol{p}_i(t)$. This effect causes a small distortion of the invariants. On the other hand, a change of velocity profile affects the invariants, because, due to its limited bandwidth, the Kalman smoother causes a distortion that depends on the frequencies contained in the motion trial. Therefore we limited the change of time scale in (13) to the range between 80 and 125 %. With this one restriction, a typical distortion due to transformation (13) is shown in Figure 4. We will make the invariants corresponding to both data sets available online.

Even though Figure 4 shows a distortion of the invariants due to transformation (13), the recognition rate drops only slightly when the experiments are repeated with the modified data set, as shown in Table I: from 95 % to 94 % and from 91 % to 86 % for the DTW-based and HMM-based approaches respectively. As a comparison, when applying the DTW approach directly to the twists, the recognition rate drops from 93 % to 46 %, see the values between brackets in table I. This illustrates the robustness of the invariant-based approach.

The DTW-based approach proved to obtain the best classification results for both the original and the transformed data set. We believe that this performance gap is mainly due to the procedure followed for the model construction. In the HMM-based approach, the construction of the model is delegated to the *Baum-Welch* approach, which is able to find only a local maximum of the likelihood function, therefore being more dependent on the initial conditions and the quality of the motion trials selected for learning. Conversely, the procedure followed for model construction in the DTW-based approach, while requiring an extra computational cost, allows to weight differently motion trials that exhibit a higher dissimilarity (expressed in terms of DTW distance). This way, the constructed models are less sensitive to outliers, and lead to better recognition rates with both data sets. The HMM approach exhibits a bigger drop of classification performance with the transformed data set, but we believe that the cause lies in the simplicity of the adopted model. We also believe that the adoption of a more complex model (*e.g.* AR-HMMs, Coupled-HMMs, see [21]) would improve the classification results. The HMM-based approach may turn out to be a better choice in case of on-line motion recognition, where computational complexity and time performance are an important factor. This extension will be the subject of future work.

Given the choice of relatively simple recognition methods, the classification results obtained support the validity of the proposed invariant representation.

We will also investigate the effect of less accurate measurement systems, with less geometric accuracy and smaller sampling frequency, on the calculated invariants and on the success rates of the recognition approaches.

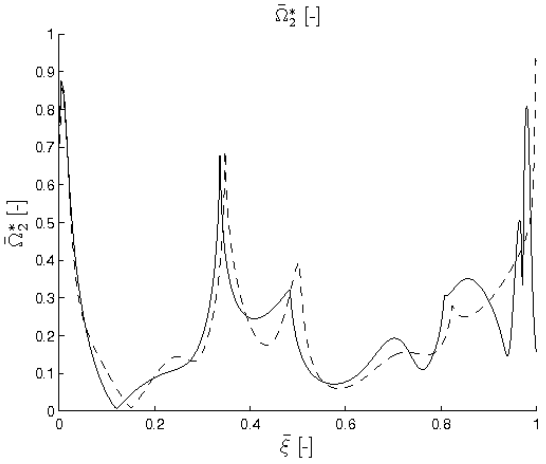

Fig. 4. Comparison of $\bar{\Omega}_2^*$ of a corresponding motion trial in data sets 1 (solid line) and 2 (dashed line).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by K.U.Leuven's Concerted Research Action GOA/2010/011 *Global real-time optimal control of autonomous robots and mechatronic systems*, the Research Council K.U.Leuven, CoE EF/05/006 *Optimization in Engineering (OPTEC)*, and European FP7 project Rosetta (FP7-230902, *Robot control for skilled execution of tasks in natural interaction with humans; based on autonomy, cumulative knowledge and learning*). Tinne De Laet is a Postdoctoral Fellow of the Fund for Scientific Research–Flanders (F.W.O.) in Belgium.

APPENDIX

This appendix contains the analytic formulas to obtain the time-based invariants starting from the twist of the rigid body, given as $\mathbf{t} = (\boldsymbol{\omega}^T \ \boldsymbol{v}^T)^T$, and its first and second time derivatives. Here $\mathbf{t}$ represents a *screw twist*, i.e. the reference point used to express the translational velocity $\boldsymbol{v}$ is the point attached to the rigid body that instantaneously coincides with the origin of the world reference frame. These formulas are derived in [3].

$$\omega_1 = \pm||\boldsymbol{\omega}|| \ ; \ v_1 = \pm\frac{\boldsymbol{v}.\boldsymbol{\omega}}{||\boldsymbol{\omega}||} \ ; \ \omega_2 = \pm\frac{||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||}{||\boldsymbol{\omega}||^2} \ ; \quad (14)$$

$$v_2 = \pm\frac{(\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}})}{||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||}\left[\frac{(\dot{\boldsymbol{\omega}} \times \boldsymbol{v} + \boldsymbol{\omega} \times \dot{\boldsymbol{v}}).||\boldsymbol{\omega}||^2 - 2(\boldsymbol{\omega} \times \boldsymbol{v}).(\boldsymbol{\omega}.\dot{\boldsymbol{\omega}})}{||\boldsymbol{\omega}||^4}\right] ; \quad (15)$$

$$\omega_3 = \pm \frac{||\boldsymbol{\omega}||}{||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||^2} . | \left(\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}} \right) . \ddot{\boldsymbol{\omega}} \ ; \tag{16}$$

$$\begin{aligned} v_3 \quad = \quad & \mp \frac{\left[\dot{\boldsymbol{\omega}} \times \left(\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}} \right) + \boldsymbol{\omega} \times \left(\boldsymbol{\omega} \times \ddot{\boldsymbol{\omega}} \right) \right] . \{ ||\boldsymbol{\omega}||^2 . (\dot{\boldsymbol{\omega}} \times \boldsymbol{v} + \boldsymbol{\omega} \times \dot{\boldsymbol{v}}) - 2 \boldsymbol{\omega} . \dot{\boldsymbol{\omega}} . (\boldsymbol{\omega} \times \boldsymbol{v}) \}}{||\boldsymbol{\omega}||^3 . ||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||^2} \\ & \mp \frac{\left(\boldsymbol{\omega} \times \left(\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}} \right) \right) . \{ ||\boldsymbol{\omega}||^2 . (\ddot{\boldsymbol{\omega}} \times \boldsymbol{v} + 2 \dot{\boldsymbol{\omega}} \times \dot{\boldsymbol{v}} + \boldsymbol{\omega} \times \ddot{\boldsymbol{v}}) - 2 (||\dot{\boldsymbol{\omega}}||^2 + \boldsymbol{\omega} . \dot{\boldsymbol{\omega}}) (\boldsymbol{\omega} \times \boldsymbol{v})) \}}{||\boldsymbol{\omega}||^3 . ||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||^2} \\ & \pm \left[\frac{3}{2} . \frac{\boldsymbol{\omega} . \dot{\boldsymbol{\omega}}}{||\boldsymbol{\omega}||^2} + \frac{(\boldsymbol{\omega} . \times \dot{\boldsymbol{\omega}}) . (\boldsymbol{\omega} \times \ddot{\boldsymbol{\omega}})}{||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||^2} \right] . \frac{\left(\boldsymbol{\omega} \times \left(\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}} \right) \right) . \{ ||\boldsymbol{\omega}||^2 . (\dot{\boldsymbol{\omega}} \times \boldsymbol{v} - \boldsymbol{\omega} \times \dot{\boldsymbol{v}}) - 2 (\boldsymbol{\omega} . \dot{\boldsymbol{\omega}}) (\boldsymbol{\omega} \times \boldsymbol{v}) \}}{||\boldsymbol{\omega}||^3 . ||\boldsymbol{\omega} \times \dot{\boldsymbol{\omega}}||^2} \ . \end{aligned} \tag{17}$$

REFERENCES

[1] Kulić, D., Takano, W., and Nakamura, Y., 2008, "Incremental learning, clustering and hierarchy formation of whole body motion patterns using adaptive Hidden Markov Chains," **27**(7), pp. 761–784.

[2] Calinon, S., Guenter, F., and Billard, A., 2007, "On learning, representing and generalizing a task in a humanoid robot," IEEE Trans. on Systems, Man and Cybernetics, Part B, **37**(2), pp. 286–298.

[3] Schutter, J. D., 2010, "Invariant description of rigid body motion trajectories," **2**(1), pp. 011004/1–9.

[4] Wu, S. and Li, Y., 2008, "On signature invariants for effective motion trajectory recognition," International J. of Robotics Research, **27**(8), pp. 895–917.

[5] Bashir, F. I., Khokhar, A. A., and Schonfeld, D., 2006, "View-invariant motion trajectory-based activity classification and recognition," Multimedia Systems, pp. DOI 10.1007/s00530–006–0024–2.

[6] Piao, Y., Hayakawa, K., and Sato, J., 2006, "Space-time invariants for recognizing 3d motions from arbitrary viewpoints under perspective projection," Trans. of the Institute of Electronics, Information and Communication Engineers, **E89-D**(7), pp. 2268–2274.

[7] Chasles, M., 1830, "Note sur les propriétés générales du système de deux corps semblables entr'eux et placés d'une manière quelconque dans l'espace; et sur le déplacement fini ou infiniment petit d'un corps solide libre," Bulletin des Sciences Mathématiques, Astronomiques, Physiques et Chimiques, **14**, pp. 321–326.

[8] Mozzi, G., 1763, *Discorso Matematico sopra il Rotamento Momentaneo dei Corpi*, Stamperia del Donato Campo, Napoli.

[9] Rauch, H., Tung, F., and Striebel, C., 1965, "Maximum likelihood estimates of linear dynamic systems," AIAA Journal, **3**(8), pp. 1445–1450.

[10] Angeles, J., 1986, "Automatic computation of the screw parameters of rigid-body motions. Part II: Infinitesimally-separated positions," **108**, pp. 39–43.

[11] Angeles, J., 1987, "Computation of rigid-body angular acceleration from point-acceleration measurements," , pp. 124–127.

[12] Angeles, J., 1988, *Rational Kinematics*, Springer.

[13] Sommer, I. H. J., 1992, "Determination of first and second order instant screw parameters from landmark trajectories," **114**, pp. 274–282.

[14] Page, A., 2009, "Experimental analysis of rigid body motion. a vector method to determine finite and infinitesimal displacements from point coordinates," **131**, pp. 031005–1–031005–8.

[15] Fenton, R. G. and Willgoss, R. A., 1990, "Comparison of methods for determining screw parameters of infinitesimal rigid body motion from position and velocity data," **112**, pp. 711–716.

[16] Sakoe, H. and Chiba, S., 1978, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Transactions on Acoustics, Speech, and Signal Processing, **26**(1), pp. 43–49.

[17] Tax, D. and Duin, R., 2008, "Growing a multi-class classifier with a reject option," Pattern Recognition Letters, **28**, pp. 1565–1570.

[18] Bishop, C. M., 2006, *Pattern Recognition and Machine Learning*, Springer.

[19] Rabiner, L. R., 1989, "A tutorial on Hidden Markov Models and selected applications in speech recognition," **77**(2), pp. 257–286.

[20] Baum, L. E., Petrie, T., Soules, G., and Weiss, N., 1970, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Annals of Mathematical Statistics, **41**, pp. 164–171.

[21] Murphy, K. P., 2002, *Dynamic Bayesian Networks: Representation, Inference and Learning*, Ph.D. thesis, UC Berkeley, Computer Science Division.